

ECON 8376 Problem Set #2 Solutions

Minwoo Yoo

George Washington University

February 5, 2026

Q1: Model Setup & Assumptions

The Model

$$y_i = g(x_i) + \epsilon_i$$

Assumptions:

① **Sampling:** The data (y_i, x_i) are i.i.d. for $i = 1, \dots, n$.

② **Error Moments:**

- $E(\epsilon_i | x_i) = 0$
- $E(\epsilon_i^2 | x_i = x) = \sigma^2(x)$ (Conditional Heteroskedasticity)

③ **Regularity Conditions:**

- Finite moments: $E(x_i)$, $E(x_i^2)$, $E(y_i)$, and $E(x_i y_i)$ are finite.
- Rank Condition: $\text{Var}(x_i) \neq 0$.

Q1 (a): OLS Consistency (1/5 - Population Target)

Question 1(a)

Argue that $\hat{\beta}_0^{ols} + \hat{\beta}_1^{ols}x$ is a consistent estimator for the best linear approximation to $g(x)$.

Step 1: Define the Linear Projection of y

The **linear projection** of y_i onto $(1, x_i)$ is:

$$L(y_i | 1, x_i) = \beta_0 + \beta_1 x_i$$

where $\beta = (\beta_0, \beta_1)'$ is defined by:

$$\beta = (E[X_i X'_i])^{-1} E[X_i y_i]$$

with $X_i = (1, x_i)'$.

We will show in Step 4 that this equals the best linear approximation to $g(x)$.

Q1 (a): OLS Consistency (2/5 - OLS Estimator)

Step 2: The OLS Estimator and Its Probability Limit

The OLS estimator is:

$$\hat{\beta}^{ols} = \left(\sum_{i=1}^n X_i X_i' \right)^{-1} \sum_{i=1}^n X_i y_i$$

Since (y_i, X_i) is i.i.d., by the **Law of Large Numbers**:

$$\frac{1}{n} \sum_{i=1}^n X_i X_i' \xrightarrow{p} E[X_i X_i']$$

$$\frac{1}{n} \sum_{i=1}^n X_i y_i \xrightarrow{p} E[X_i y_i]$$

Therefore: $\hat{\beta}^{ols} \xrightarrow{p} \beta = (E[X_i X_i'])^{-1} E[X_i y_i]$

Q1 (a): OLS Consistency (3/5 – Regularity Conditions: Why Needed?)

We need to check some **conditions** that ensure:

- the population target β is **well-defined**
- the probability limit of $\hat{\beta}^{OLS}$ **exists**

For

$$\beta = (E[X_i X'_i])^{-1} E[X_i y_i]$$

to exist, we require:

- ① $E[X_i X'_i]$ is **finite**
- ② $E[X_i X'_i]$ is **invertible**
- ③ $E[X_i y_i]$ is **finite**

Q1 (a): OLS Consistency (4/5 – Verifying the Conditions)

Step 3: Verifying Regularity Conditions

Matrix:

$$E[X_i X'_i] = E\left[\begin{pmatrix} 1 \\ x_i \end{pmatrix} (1 \quad x_i)\right] = \begin{pmatrix} 1 & E(x_i) \\ E(x_i) & E(x_i^2) \end{pmatrix}$$

Condition 1 (Finite): $E(x_i)$ and $E(x_i^2)$ are finite by assumption. ✓

Condition 2 (Invertible):

$$\det(E[X_i X'_i]) = E(x_i^2) - [E(x_i)]^2 = \text{Var}(x_i)$$

Since $\text{Var}(x_i) \neq 0$, the matrix is invertible. ✓

Condition 3 (Finite $E[X_i y_i]$): Components $E(y_i)$ and $E(x_i y_i)$ are finite by assumption. ✓

Conclusion: β is well-defined, so $\hat{\beta}^{OLS} \xrightarrow{P} \beta$.

Q1 (a): OLS Consistency (5/5- From y to $g(x)$)

Step 4: Why Best Linear Predictor of y = Best Linear Approximation to $g(x)$?

Equivalence: Substitute $y_i = g(x_i) + \epsilon_i$:

$$E[(y_i - X'_i\beta)^2] = E[(g(x_i) - X'_i\beta)^2] + 2E[(g(x_i) - X'_i\beta)\epsilon_i] + E[\epsilon_i^2]$$

By LIE, the cross-term vanishes:

$$E[(g(x_i) - X'_i\beta)\epsilon_i] = E_x \left[(g(x_i) - X'_i\beta) \underbrace{E[\epsilon_i | x_i]}_{=0} \right] = 0$$

Conclusion: Minimizing MSE for y is equivalent to minimizing MSE for $g(x)$.

Thus, $\hat{\beta}^{ols} \xrightarrow{P} \beta$, the best linear approximation to $g(x)$.

Q1 (b): Infeasible WLS (1/7 - Setup)

Question 1(b)

Consider the estimator that minimizes $\sum_{i=1}^n \sigma^{-2}(x_i)(y_i - b_0 - b_1 x_i)^2$.
Is it consistent for the best linear approximation to $g(x)$?

Step 1: Defining the Sample Weights

Define sample weights:

$$\hat{w}_i = \frac{\sigma^{-2}(x_i)}{n^{-1} \sum_{j=1}^n \sigma^{-2}(x_j)}$$

Note: $\sum_{i=1}^n \hat{w}_i = n$, so $\frac{1}{n} \sum_{i=1}^n \hat{w}_i = 1$.

Define weighted means:

$$\bar{x}_w = \frac{1}{n} \sum_{i=1}^n \hat{w}_i x_i \quad \text{and} \quad \bar{y}_w = \frac{1}{n} \sum_{i=1}^n \hat{w}_i y_i$$

Q1 (b): Infeasible WLS (2/7 - Deriving the Estimator)

Step 2: Deriving the WLS Estimator from FOC

Taking FOC and solving, the WLS estimators can be written as:

$$\hat{\beta}_1^{wls} = \frac{\sum_{i=1}^n \hat{w}_i(x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum_{i=1}^n \hat{w}_i(x_i - \bar{x}_w)^2}$$

$$\hat{\beta}_0^{wls} = \bar{y}_w - \hat{\beta}_1^{wls} \bar{x}_w$$

Q1 (b): Infeasible WLS (3/7 - Population Weight)

Step 3: Applying LLN to Find the Probability Limit

(i) Define the **population weight**:

$$w_i = \frac{\sigma^{-2}(x_i)}{E[\sigma^{-2}(x_i)]} \quad \text{with} \quad E[w_i] = 1$$

(ii) By LLN, $\hat{w}_i \xrightarrow{P} w_i$ since the denominator converges:

$$n^{-1} \sum_{j=1}^n \sigma^{-2}(x_j) \xrightarrow{P} E[\sigma^{-2}(x_i)]$$

(iii) Define weighted moments (using population weight w_i):

$$E_w(x_i) = E[w_i x_i], \quad E_w(y_i) = E[w_i y_i]$$

$$\text{Var}_w(x_i) = E[w_i(x_i - E_w(x_i))^2]$$

$$\text{Cov}_w(x_i, y_i) = E[w_i(x_i - E_w(x_i))(y_i - E_w(y_i))]$$

Q1 (b): Infeasible WLS (4/7 - Probability Limit)

Step 3 (cont'd): The Probability Limit

(iv) By LLN, the WLS estimators converge to:

$$\beta_1^{wls} = \frac{Cov_w(x_i, y_i)}{Var_w(x_i)}, \quad \beta_0^{wls} = E_w(y_i) - \beta_1^{wls} E_w(x_i)$$

Key Point: This is **different** from the OLS limit β in general!

The weighted moments E_w , Var_w , Cov_w differ from the unweighted ones.

Q1 (b): Infeasible WLS (5/7 - Population Interpretation)

Step 4: What Does β^{wls} Solve?

Consider the **weighted MSE** minimization problem:

$$\min_{b_0, b_1} E \left[w_i (y_i - b_0 - b_1 x_i)^2 \right]$$

Taking FOC gives the solution:

$$b_1^* = \frac{Cov_w(x_i, y_i)}{Var_w(x_i)}, \quad b_0^* = E_w(y_i) - b_1^* E_w(x_i)$$

This is exactly β^{wls} from Step 3! So β^{wls} solves the weighted MSE problem.

By the same logic as Q1(a) (cross-term vanishes by LIE), this is equivalent to:

$$\min_{b_0, b_1} E \left[w_i (g(x_i) - b_0 - b_1 x_i)^2 \right]$$

This is **different from** β in general!

β^{wls} minimizes **weighted MSE**, not unweighted MSE.

Q1 (b): Infeasible WLS (6/7 - Density Interpretation)

Step 5: Reinterpreting as a Different Distribution

If x_i has density $f_x(x)$, then:

$$E[w_i(g(x_i) - b_0 - b_1x_i)^2] = \int w(x)(g(x) - b_0 - b_1x)^2 f_x(x) dx$$

where $w(x) = \sigma^{-2}(x)/E[\sigma^{-2}(x_i)]$.

Define $f^*(x) = w(x)f_x(x)$. Then:

$$\int f^*(x) dx = \int w(x)f_x(x) dx = E[w_i] = 1$$

So $f^*(x)$ is a valid probability density!

The weighted MSE becomes:

$$\int (g(x) - b_0 - b_1x)^2 f^*(x) dx$$

Q1 (b): Infeasible WLS (7/7 - Conclusion)

Step 6: Final Answer

Conclusion

$\hat{\beta}^{wls}$ is **NOT** consistent for the best linear approximation to $g(x)$.

Instead, $\hat{\beta}^{wls}$ is consistent for:

- The best linear approximation to $g(x)$ under the **distorted density** $f^*(x) = w(x)f_x(x)$
- Equivalently: minimizes the **weighted MSE** with weights $\sigma^{-2}(x)$

Exception: If $\sigma^2(x)$ is constant (homoskedasticity), then $w(x) = 1$ and $\beta^{wls} = \beta$.

Q1 (c): Can OLS and WLS Have Opposite Signs? (1/2)

Question 1(c)

Is it possible to have $\text{plim}_n \hat{\beta}_1^{\text{ols}} > 0$ and $\text{plim}_n \hat{\beta}_1^{\text{wls}} < 0$?

Intuition:

Consider the case where $\sigma^2(x)$ is **increasing** in x :

- Then $\sigma^{-2}(x)$ is **decreasing** in x
- The weights $w(x)$ give **more weight to smaller x**
- The distorted distribution $f^*(x)$ is **shifted to the left**

Effect: WLS effectively fits the line on a left-shifted version of the data.

Q1 (c): Can OLS and WLS Have Opposite Signs? (2/2)

When Can Signs Differ?

- If $g(x)$ is **monotonic** (e.g., increasing), both β_1 and β_1^{wls} will typically have the same sign
- Magnitude may differ: if $\sigma^2(x)$ is increasing, typically $\beta_1^{wls} < \beta_1$

Sign Reversal is Possible but Extreme:

- Requires $g(x)$ to be **non-monotonic**
- Example: $g'(x) < 0$ for small x , $g'(x) > 0$ for large x
- If weights shift distribution left enough, β_1^{wls} could be negative while $\beta_1 > 0$

Answer: Yes, it is possible, but would require a fairly extreme scenario.

Q2: MDS, Stationarity, and Ergodicity - Setup

Model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \text{ with } E(x_i \epsilon_i) = 0 \text{ and } (y_i, x_i) \text{ i.i.d.}$$

Key Definitions:

- **Martingale Difference Sequence (MDS):** $E(z_i | z_{i-1}, \dots, z_1) = 0$
- **Stationary:** Joint distribution of $(z_{i_1}, \dots, z_{i_r})$ same as $(z_{i_1+j}, \dots, z_{i_r+j})$
- **Ergodic:** $\lim_{n \rightarrow \infty} E[f(z_i, \dots, z_{i+k})g(z_{i+n}, \dots, z_{i+k+n})] = E[f(z_i, \dots, z_{i+k})]E[g(z_{i+n}, \dots, z_{i+k+n})]$

Q2 (a): $x_i\epsilon_i$ is Stationary, Ergodic MDS (1/2)

Question 2(a)

Show that $x_i\epsilon_i$ is a stationary and ergodic martingale difference sequence.

Step 1: $x_i\epsilon_i$ is i.i.d.

- (y_i, x_i) is i.i.d. by assumption
- $\epsilon_i = y_i - \beta_0 - \beta_1 x_i$ is a function of (y_i, x_i)
- Therefore (ϵ_i, x_i) is i.i.d. (change of variables)
- Hence $z_i = x_i\epsilon_i$ is i.i.d.

Q2 (a): $x_i\epsilon_i$ is Stationary, Ergodic MDS (2/2)

Step 2: Stationarity

Since $z_i = x_i\epsilon_i$ is i.i.d., it has the same distribution F_z for all i .

- $(z_{i_1}, \dots, z_{i_r})$ has distribution $\prod_{k=1}^r F_z$
- Same as $(z_{i_1+j}, \dots, z_{i_r+j})$ for any j ✓

Step 3: Ergodicity

By independence (i.i.d.), for all $n > k$:

$$E[f(z_i, \dots, z_{i+k})g(z_{i+n}, \dots, z_{i+n+k})] = E[f(\cdot)]E[g(\cdot)]$$

Hence the limit as $n \rightarrow \infty$ trivially holds. ✓

Step 4: MDS

By independence: $E(z_i | z_{i-1}, \dots, z_1) = E(z_i) = E(x_i\epsilon_i) = 0$ ✓

Q2 (b): $\epsilon_i \epsilon_{i-1}$ is MDS

Question 2(b)

Suppose ϵ_i is i.i.d. **with** $E(\epsilon_i) = 0$. Show that $z_i = \epsilon_i \epsilon_{i-1}$ is a MDS.

Proof: Let $z_i = \epsilon_i \epsilon_{i-1}$. Note: $\sigma(z_{i-1}, \dots, z_1) \subset \sigma(\epsilon_{i-1}, \dots, \epsilon_0)$.

$$\begin{aligned}E(z_i | z_{i-1}, \dots, z_1) &= E [E(\epsilon_i \epsilon_{i-1} | \epsilon_{i-1}, \dots, \epsilon_0) | z_{i-1}, \dots, z_1] \quad (\text{LIE}) \\&= E [\epsilon_{i-1} \cdot E(\epsilon_i | \epsilon_{i-1}, \dots, \epsilon_0) | z_{i-1}, \dots, z_1] \quad (\epsilon_{i-1} \text{ measurable}) \\&= E [\epsilon_{i-1} \cdot E(\epsilon_i) | z_{i-1}, \dots, z_1] = 0 \quad (\text{i.i.d.} + E(\epsilon_i) = 0)\end{aligned}$$

Mathematical Foundation: Conditional Expectation

The Fundamental Identity

Let (Ω, \mathcal{F}, P) be a probability space. For an integrable R.V. Y and a sub- σ -field $\mathcal{G} \subset \mathcal{F}$, the conditional expectation $E[Y|\mathcal{G}]$ is the unique \mathcal{G} -measurable function such that:

$$\int_{\omega \in A} E[Y|\mathcal{G}](\omega) dP(\omega) = \int_{\omega \in A} Y(\omega) dP(\omega) \quad \forall A \in \mathcal{G}$$

Set-Theoretic Hierarchy:

- **Integration Domain (A):** The integral is performed over the set of outcomes $\omega \in A$.
- **Information Set (\mathcal{G}):** The σ -field \mathcal{G} is the *collection* of all measurable events A . It dictates which "subsets of Ω " we are allowed to integrate over.
- **Identity Meaning:** The summary $E[Y|\mathcal{G}]$ must preserve the total probability mass of Y for every measurable event A defined by the information \mathcal{G} .

Formal Proof: The Tower Property

Theorem: Let $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$. Then $E[E[Y|\mathcal{H}]|\mathcal{G}] = E[Y|\mathcal{G}]$.

Proof Strategy: Show that for any $A \in \mathcal{G}$, the integral of both sides over A is identical.

- ① By the definition of the **outer** expectation $E[\cdot|\mathcal{G}]$:

$$\int_{\omega \in A} E[E[Y|\mathcal{H}]|\mathcal{G}] dP = \int_{\omega \in A} E[Y|\mathcal{H}] dP \quad (\text{since } A \in \mathcal{G})$$

- ② Since $\mathcal{G} \subset \mathcal{H}$, any set A in our list \mathcal{G} is also in the expert's list \mathcal{H} ($A \in \mathcal{H}$). Applying the definition of the **inner** expectation $E[Y|\mathcal{H}]$:

$$\int_{\omega \in A} E[Y|\mathcal{H}] dP = \int_{\omega \in A} Y dP \quad (\text{since } A \in \mathcal{H})$$

Conclusion: Because $\int_A E[E[Y|\mathcal{H}]|\mathcal{G}] dP = \int_A Y dP$ for all $A \in \mathcal{G}$, it follows that $E[E[Y|\mathcal{H}]|\mathcal{G}]$ satisfies the definition of $E[Y|\mathcal{G}]$.

Q2 (c): Variance of Sample Mean (1/2)

Question 2(c)

Given $\text{Var}(z_i) = \sigma^2$, $\text{Cov}(z_i, z_j) = 0$ if $|i - j| > n/2$, and $\text{Cov}(z_i, z_j) = c \neq 0$ if $|i - j| \leq n/2$.
Find $\lim_n \text{Var}(\bar{z})$.

Setup:

$$\begin{aligned}\text{Var}(\bar{z}) &= \text{Var} \left(\frac{1}{n} \sum_{i=1}^n z_i \right) = \frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n z_i \right) \\ &= \frac{1}{n^2} \left[n\sigma^2 + 2 \sum_{i < j} \text{Cov}(z_i, z_j) \right]\end{aligned}$$

Q2 (c): Variance of Sample Mean (2/3)

Counting Pairs: For gap $k = j - i$, since $j = i + k \leq n$, we have $i \leq n - k$.
So $i \in \{1, 2, \dots, n - k\} \Rightarrow (n - k)$ pairs for each k .

Case 1: n even (e.g., $n = 6 \Rightarrow k = 1, 2, 3$)

$$\sum_{k=1}^{n/2} (n - k) = \frac{n}{2} \times \frac{(n-1) + \frac{n}{2}}{2} = \frac{3n^2}{8} + O(n)$$

Case 2: n odd (e.g., $n = 5$: condition $k \leq n/2 = 2.5$, but $k \in \mathbb{Z}^+ \Rightarrow k = 1, 2$)

$$\sum_{k=1}^{(n-1)/2} (n - k) = \frac{n-1}{2} \times \frac{(n-1) + \frac{n+1}{2}}{2} = \frac{3n^2}{8} + O(n)$$

Q2 (c): Variance of Sample Mean (3/3)

Final Calculation:

$$\begin{aligned} \text{Var}(\bar{z}) &= \frac{1}{n^2} \left[n\sigma^2 + \frac{3n^2}{4}c + O(n) \right] \\ &= \underbrace{\frac{\sigma^2}{n}}_{\rightarrow 0} + \frac{3c}{4} + \underbrace{O(1/n)}_{\rightarrow 0} \end{aligned}$$

Result:

$$\lim_{n \rightarrow \infty} \text{Var}(\bar{z}) = \frac{3c}{4}$$

Key Insight: The variance does **not** vanish!

The n -dependent covariance structure contributes $O(n^2)$ terms that survive after dividing by n^2 .

Q3: Monte Carlo Simulation - Results

Question 3

Monte Carlo study of HC standard errors and rejection rates (Size).

Simulation Results vs. Benchmark:

	n = 10		n = 20		n = 50	
	Mean	SE	Size	Mean	SE	Size
True SD	0.4591		–	0.3122		–
HC0	0.2314	0.3467	0.1824	0.3081	0.1339	0.2308
HC1	0.2587	0.3103	0.1922	0.2886	0.1367	0.2223
HC2	0.3246	0.2103	0.2343	0.2046	0.1586	0.1633
HC3	0.5339	0.0962	0.3309	0.1174	0.1977	0.1077

- Benchmark:** The "True SD" is the actual empirical SD of $\hat{\beta}_1$ across 10,000 simulations.
- Downward Bias:** HC0 ~ HC2 are much smaller than the True SD (Biased Downward).

Q3: Monte Carlo - Why Rejection Rates Are Too High

The Puzzle:

HC3 is nearly unbiased (e.g., at $n = 50$, $0.1977 \approx 0.1957$), yet the rejection rate (0.1077) is still double the nominal size (0.05). Why?

Answer: Variability in Standard Errors

- The estimated standard error \widehat{SE} is a **random variable**, not a constant.
- In small samples, \widehat{SE} has high variance (it fluctuates a lot).
- When \widehat{SE} happens to be **small** (by chance), the t-statistic $t = \hat{\beta}/\widehat{SE}$ becomes **artificially large**.
- This "noise" in the denominator creates heavier tails than the standard normal/t-distribution expects, causing **over-rejection**.

Q4: Bootstrap - Results

Question 4

Bootstrap inference for a nonlinear function of parameters.

Part (a): Point Estimate and CI

- Estimate: $\hat{\theta} = 2.713$
- 95% CI (nlcom): [1.002, 4.225]
- 90% CI (nlcom): [1.277, 4.150]

Part (b): Bootstrap Standard Error

- Bootstrap SE: 61.41 (or 3.82 in another run)
- nlcom SE: 0.87
- **Problem:** Outliers in bootstrap distribution cause instability

Q4: Bootstrap - Percentile Interval

Part (c): Bootstrap Percentile Interval

- Bootstrap 95% CI: [1.495, 8.020] (or [1.601, 7.970])
- This is **wider** than the nlcom interval
- **Not centered** around the estimate of 2.713

Key Insight:

- Bootstrap SE is **unstable** due to outliers
- Bootstrap **percentile interval** is more robust
- Percentile interval captures asymmetry in the distribution
- For nonlinear functions, bootstrap can reveal non-normality